DOCUMENT RESUME

ED 150 631 CS 203 958

AUTHOR Cronnell, Bruce; Rhode, Mary
TITLE Linguistic Information in the Vocabulary of Elementary School Children.
PUB DATE Nov 77
NOTE 12p.; Paper presented at the Annual Meeting of the National Council of Teachers of English (67th, New York City, November 24-26, 1977) ; Document may not reproduce clearly due to print quality

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Child Language; Codification; *Communication Skills; Elementary Education; Language Research; Language Skills; *Lexicography; *Linguistics; *Vocabulary; Word Frequency; *Word Lists

ABSTRACT

This paper discusses the procedures employed in compiling and coding a basic set of vocabulary words and related linguistic data intended for use in elementary communication skills instruction. Ten thousand words were selected from eight studies of words used by children at various grade levels, one through six. Each word was assigned a grade level and coded with a variety of information retrievable by computer. Separate codes were used to indicate syllables and stress, compound words, spelling-sound correspondences, affixes, affixation processes, and word bases. Tables on word frequency and the rank order of prefixes and suffixes are included. (HAI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED150631

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Bruce Cronnell
Mary Rhode

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper to be presented
at NCTE convention,
November 1977

LINGUISTIC INFORMATION IN THE VOCABULARY OF ELEMENTARY SCHOOL CHILDREN

Bruce Cronnell and Mary Rhode
SWRL Educational and Research Laboratory
4665 Lampson Avenue
Los Alamitos, CA 90720

The goal of this lexical study was to provide a basic set of words with accompanying linguistic data for use in the development of elementary communication skills instruction. This paper summarizes the procedures undertaken to compile and code the lexicon and the results of this compilation and coding.

Why another lexicon?

Many word lists are presently available so it may be wondered why another lexicon is needed. There are several reasons:

1. Many word lists are based on professionally written text. That is, they are studies of words that are read in books. We were concerned with the words actually used by children, whether or not they encounter them in their reading.
2. Many of the studies available are directed toward adult use or toward use by students in grades other than K-6, which was our concern.
3. Many of the studies are old. Since vocabulary is the fastest-changing part of language, word lists must be constantly updated.

Note: This paper is written in a conversational manner to represent the way that the presentation will be given. Editing would be necessary before publication.

CS203958

But a primary reason for this word list was that we wanted linguistic information about the words used by children. This information can be used for research and development purposes and is not available in any other study of vocabulary.

Where did the words come from?

The ideal way of obtaining a lexicon of words used by children would be to collect large samples of children's speech and writing. Since we had neither the time nor the money to undertake such a large-scale study, we depended on the work already done.

Thus we compiled words from eight different studies of words used by children. The studies were from five to 30 years old. They covered children at various grade levels--from a single grade to six grades. They included both words spoken by children and words written by children, as well as a limited set of words used in elementary school texts.

Since the studies we used varied in the ways that they treated proper nouns, we undertook separate studies of personal names and of geographic names. These studies provided nearly 400 personal names and about 250 geographic names.

Because very few large-scale vocabulary-use studies have been conducted in recent years, we also conducted a study to obtain words that have come into the language in the past 30 years. Using six published lists of new words, we obtained over 300 new words known by elementary school children, but not found in our other sources.

From all this compilation, we gathered a lexicon of over 10,000 words used by elementary school children.

(I should note here that by "word" we mean a base--or root--word; that is a word without inflectional suffixes like -s, -ed, -ing.)

How were the words coded?

The word list in and of itself was not our primary concern. Rather we were interested in information about these words. Thus we coded each word with a variety of information that could be retrieved by computer. Information about each word was coded on two computer cards, thus providing over 20,000 lines of information. I'll describe some of the coding together with the results of the coding. As you can imagine, with so much information available, I can report only a small part of the results of our analyses. In addition, we have only touched the surface in our analyses so far; much more work remains to be done by us and others.

Grade level

Based on information available in our sources, we assigned a grade level of K to 6 to each word. (For some words--primarily the new words--we couldn't assign a grade level; these are ungraded words.) Part A on the handout indicates frequency by grade level. It's interesting to note that fully a quarter of the words are known by kindergarteners. In fact, we suspect that even more words are known at the kindergarten level, but we were conservative in our assignment of grade level. A study of nearly 10% of the lexicon indicated that words were often known by students at a grade level lower than that assigned. You will note that a large number of words also occur at the sixth-grade level. This is due to a "ceiling effect": since our cut-off point was sixth grade, words tended to bunch at that level.

Frequency

We assigned each word a number to indicate its relative frequency of use in textbooks. This was based on the information in the Carroll, Davies, and Richman Word frequency book. Part B of the handout shows a graph of the frequency distribution, with 1 indicating low frequency and 7 indicating high frequency. There are relatively few high-frequency words--as would be expected, since all studies of word frequency find a small set of common words used with very great frequency. And we were pleased to find that very few of our words occurred at low frequencies. The many words at medium frequencies are those most useful for communicating about a variety of topics.

Syllables and Stress

Each word was coded for the number of syllables in it. Over a quarter of the words were one-syllable words. Nearly half were two-syllable words. A fifth were three-syllable words, with less than 10% longer than three syllables.

Each word was also coded for the pattern of stressed (accented) and unstressed (unaccented) syllables. One-syllable words, of course, have no pattern. In the two- and three-syllable words, the predominant patterns included stress on the first syllable. Over 25% of the words in the lexicon were two-syllable words with the first syllable stressed and the second syllable unstressed.

Compounds

Over 10% of the lexicon is composed of compound words: two words written as one with either no space or a hyphen between them. An analysis of these compounds indicates that a majority of them can be organized

into one of 14 patterns that relate their meaning to the use of the individual words in a sentence. For example, when someone runs away, that person is a runaway; a bluebird is a bird that is blue; when your ear aches, you have an earache.

Spelling-Sound Correspondences

Each word was coded for the spelling-sound correspondences found in it. This information can be used for phonics instruction in both reading and spelling.

We recognized 76 different spelling units, that is, letters or letter sequences, like i, n, nn, ea, sh. Most of these spelling units represent more than one pronunciation, so they were marked with numerical codes, often supplemented with other symbols to indicate additional features of interest.

Some examples of this coding are shown in Part C on the handout. In wink, wine, and wire, the initial w is coded "W" since the normal correspondence for this letter is /w/. In wink, the i is coded "115" to represent the /ɪ/ sound. In wine and wire, the i is coded "111" to represent the /ay/ sound and the final e is coded "E18" to represent the final silent e marker. The "+" in wire indicates that the vowel is followed by r, which tends to affect the pronunciation of the vowel. Since /k/ is the most frequent pronunciation of k, no code is added to the "K." Since r generally has only one pronunciation, no extra code is added to the "R." However, n has two pronunciations. The most common one, as in wine, is coded with a plain "N" while the one corresponding to /ŋ/, as in wink, is coded "N4."

This coding system is somewhat complex, but it permits recognition of a great number of different spelling arrangements, although many generalizations can be made across these differences. Of course, the large number of codes is far more than would be used in actual reading or spelling instruction. However, the great detail is essential for use by researchers and by curriculum designers.

Affixes

Affixes--prefixes and suffixes--are very common in words of more than one syllable. We were very liberal in our definition of affixes, including such forms as the ob- in object and the -om in bottom. With this liberal definition, we found that nearly 60% of the words in the lexicon were affixed, including many that had more than one affix. In fact, there were over 8000 occurrences of affixes in the lexicon. Even using a more conservative approach, we found about 3500 affixes that were attached to independent words. Clearly affixes are an important part of the vocabulary used by elementary school students.

We identified 34 different prefixes occurring more than five times each, and 84 suffixes occurring more than five times each. All but 12 of these suffixes occur with independent words. A rank-ordered list of all these affixes is found in Parts D and E of the handout.

Many of these affixes are subject to spelling and pronunciation changes. Take, for example, the prefix con-, as illustrated in Part F on the handout. (Con- has more variation than most affixes, but is a useful example to show the range of variations.) It is pronounced /kan/ in concept, /kaŋ/ in congress, or /kən/ in conserve. In addition, the n of con- may

assimilate with the following consonant, often with changes in the pronunciation of the vowel. For example, it is pronounced /kam/ in combat, /kʌm/ in company, /kəl/ in college, /kəl/ in collect, /kor/ in correspond, and /kər/ in correct. These spelling and pronunciation variations are individually coded to permit detailed analysis of word-formation processes. However, as with spelling-sound correspondences, such detail might not be appropriate for instructional purposes.

Affixation Processes

When affixes are added to independent words, changes frequently take place in the base word. We identified nearly 20 different changes. Some of these changes involve only spelling variations; for example, e dropping and final consonant doubling. Others, while involving some minor spelling change, are notable for a change in pronunciation that is not marked in the spelling; for example, the stress and vowel changing from reside to resident. Still others involve notable spelling and pronunciation changes together; for example, le (/əl/) changing to il (/ɪl/), as in able-ability, or t (/t/) changing to c (/s/), as in prophet-prophecy.

Such affixation processes occurred about 2000 times, although in far fewer words, since more than one of the processes could occur in a single word. However, it should be noted that in text more affixation processes might be expected to be present in inflected words, which were not included in this study.

Bases

Affixes--prefixes and suffixes--are added to bases. For example the prefix re- can be added to the bases write and ject to form rewrite and

reject. We have been particularly interested in bound bases like ject, which derive from Latin or Greek. Bound bases cannot stand as independent words, but must be affixed--generally prefixed. We identified nearly 200 such bound bases that occurred with some frequency in the lexicon. For example, the base tract which generally means "pull," occurs in tractor, attract, contract, extract, and subtract. And, of course, many of these words can be affixed further; for example, attractive, contractor, extraction, subtraction. Thus the combination of bases and affixes is seen to be a useful and productive word-building process.

Conclusion

A period of several years was spent compiling and coding this communication skills lexicon. While some of the words may not be entirely appropriate and while some appropriate words are undoubtedly missing, this vocabulary base is large enough to provide an adequate representation of the words used by elementary school children.

We attempted to be careful and consistent in our coding, but with nearly 200,000 individual codes used to cover the 10,000 words, the chance of error is always there. In addition, the codes themselves may be seen as inappropriate in light of other views of language. However, the codes provide an extensive and detailed indication of the complexity and potential of children's vocabulary. They can thus be of great value in future research and development activities at the elementary school level.

Reference

Complete documentation of this study is found in the following report:

Rhode, M. and Cronnell, B. Compilation of a communication skills lexicon coded with linguistic information. Technical Report No. 58, 1977, SWRL Educational Research and Development, Los Alamitos, California.

NCTE 1977

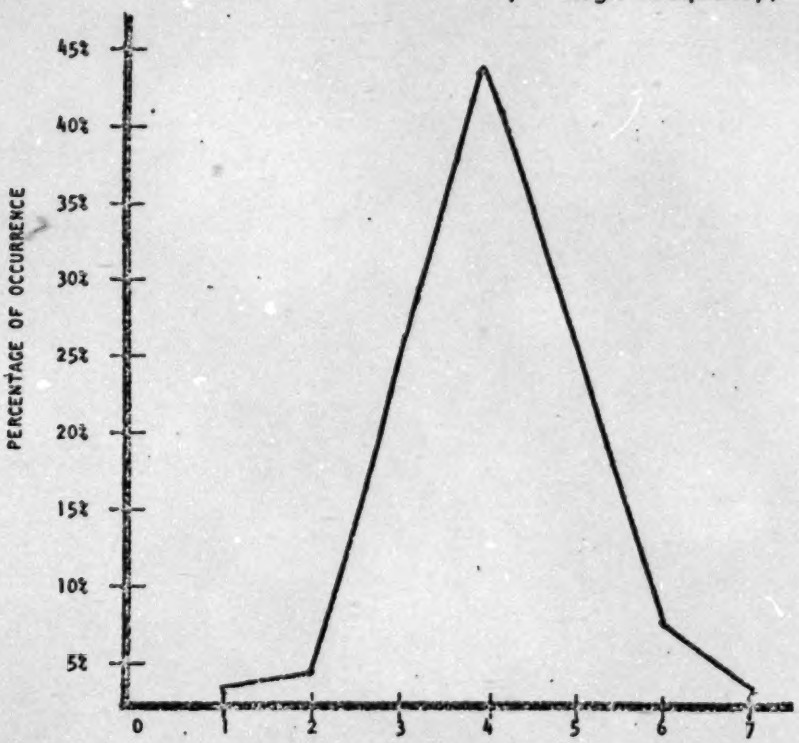
LINGUISTIC INFORMATION IN THE VOCABULARY OF ELEMENTARY SCHOOL CHILDREN

Bruce Cronnell and Mary Rhode
SWRL Educational Research and Development
4665 Lampson Avenue
Los Alamitos, CA 90720

A. Grade Level Frequencies

Grade Level	Frequency	% of Lexicon
K	2496	25%
1	1225	12%
2	1033	10%
3	987	10%
4	904	9%
5	1126	11%
6	2055	20%
Ungraded	302	3%
Total	10,128	100%

B. Occurrence of Low to High Frequency Words (1 = low frequency; 7 = high frequency)



C. Sample Coding for Spelling-Sound Correspondences

wink	W,I15,N4,K
wine	W,I11N,E18
wire	W,I11+,R,E18

D. Rank Order of Prefixes

Prefixes	Frequency Totals
CON	249
RE	200
AD	172
IN	134
DE	122
EX	100
DIS	94
A	86
UN	76
PRO	72
EN	68
E	55
PRE	44
SUB	41
BE	38
OB	38
IN#	35
SE	29
PER	22
INTER	21
AB	17
TRANS	15
SUR	13
FOR	12
PARA	11
SUPER	10
SYN	10
MIS	9
TELE	8
DIA	7
AL	6
ANTI	6
MONO	5
MULTI	5

E. Rank Order of Suffixes

Suffix	Frequency Totals	Suffix	Frequency Totals
ER	945	UM	27
Y	472	US	27
ING	367	LESS	25
ION	302	ORY	24
LE	300	ERN	23
ED	287	ID	23
LY	260	IL	23
ATE	218	WARD	23
AL	204	ICE	22
S	191	IFY	22
EN	154	IZE	22
OR	146	ARD	21
IC	128	UTE	21
ON	106	AIN	19
ENT	83	ITE	19
MENT	83	OM	19
ET	82	ILE	18
OUS	78	IST	17
EL	74	EST	16
AN	65	IBLE	16
IN	62	TEEN	14
URE	59	SHIP	13
FUL	57	A	12
ISH	54	OT	12
AGE	53	RY	12
ITY	53	ADE	11
INE	51	ESS	11
AR	48	IO	11
IVE	47	EM	9
ANT	46	UNE	9
ABLE	41	ETTE	8
ARY	40	ISM	8
IA	39	LET	7
ENCE	38	T	7
IE	38	UDE	7
TH	36	OP	6
ANCE	35	ACLE	5
ERY	31	EE	5
IS	31	HOOD	5
IT	30	SOME	5
NESS	28	STER	5
TY	27	ULE	5

F. Examples of Variations in Prefix Spelling and Pronunciation: con-

concept	Assimilated:	combat	company
congress		college	collect
conserve		correspond	correct

G. Reference

Rhode, M. and Cronnell, B. Compilation of a communication skills lexicon coded with linguistic information. Technical Report No. 58, 1977, SWRL Educational Research and Development, Los Alamitos, California.